INVESTIGATIONS ON CARTOGRAPHIC CONSTRAINT FORMALISATION

Dirk Burghardt[1], Stefan Schmid[1] and Jantien Stoter[2]

[1]{burg, sschmid}@geo.uzh.ch
Department of Geography, University Zurich
Winterthurerstrasse 190, CH-8057 Zurich
[2]stoter@itc.nl
International Institute for Geo-Information Science and Earth Observation
P.O. Box 6, 7500 AA Enschede, the Netherlands

Abstract: This paper presents investigations on the formalisation of cartographic requirements by constraint specifications. The work was carried out within the EuroSDR[1] project on studying the state-of-the-art of commercially available generalisation software. The aim of the study presented in this paper was to analyse and compare several constraint specifications from various National Mapping Agencies (ICC, Catalonia; IGN, France; OS, Great Britain; TD Kadaster, Netherlands)[2] which will be the input for the EuroSDR tests. In a first step the NMA prepared small geo-data sets, which contained characteristic map situations such as urban, mountainous, rural and costal area. In addition they expressed cartographic requirements of the target map with help of a template for constraint specification, which was provided by the EuroSDR project team. This paper describes how the NMAs specified their constraints. Based on the constraints analysis, the cartographic constraints were first harmonised in a generic set of constraints that can be used by all four NMAs within this project. This harmonisation of constraints is presented in the paper.
Apart from the harmonisation, a result of the study on the constraints is the definition of an extended typology of cartographic constraints which is also presented. A frequency and granularity of constraint specifications is analysed against this topology and graphically presented. The influence of data model for constraint specification is shown. Finally difficulties and open questions for the constraint specification are discussed.
After these investigations there exists not a complete set of cartographic constraints, because the specified constraints refer only to the selected examples. But as a result, the harmonisation of frequent defined constraints was reached. The evaluation process within the EuroSDR project will show if the specified constraints are sufficient for the description of cartographic requirements of the target map.

1. Introduction

Past research on improving generalisation of topographic data has shown that generalisation is not an easy task. Individual geometrical transformations triggered by generalisation requirements will often impact upon each other. This interdependency between operators asks for optimisation techniques designing the optimal sequence of individual operators, applied as algorithms with appropriate parameter values. In this paper it is assumed that for this optimisation one should not concentrate on implementing the most optimal generalisation process, i.e. trying in advance to determine the best order of operators including the algorithms and parameter values. Instead the starting point of generalisation aiming at readable cartographic output should be a precise description of what one wants to achieve by the generalisation process. This precise description of generalisation output can be accomplished by a formal specification of cartographic requirements of the output. In an iterative process the system can then analyse which combination of operators (with their algorithms and operators) gives the best overall result.

---

[1] EuroSDR - European Spatial Data Research Organisation (formerly OEEPE)
[2] National Mapping Agencies (NMA):
ICC - Institut Cartogràfic de Catalunya; IGN - Institut Géographique National, OS - Ordnance Survey; TD Kadaster – Topografische Dienst Kadaster

A good way to specify cartographic output is to define the output as a set of constraints. This will be the focus of this paper. Research on achieving satisfying generalisation output based on cartographic constraints specification has yielded promising results. But how does cartographic constraints specification work in practice?

This paper reports on findings of the EuroSDR project that studies the state-of-the-art of commercially available generalisation software (EuroSDR, 2007). The project is executed by members from several NMAs as well as from research institutes (see acknowledgement). Main objective of the project is generalisation of complete topographic data sets at scales larger than 1:50k. The generalisation aims at obtaining a digital cartographic model from a digital landscape model. Four case studies were defined covering four areas in four different countries to study commercially available generalisation software. The assumption of the project team is that that specification of cartographic constraints in a formal way allows designing the most optimal generalisation scenario. In addition a formal description of constraints provides a objective method to evaluate different generalisation solutions, i.e. the one that leads to the most satisfied constraints will be evaluated as the best solution. In the preparation phase of the EuroSDR project, started in October 2006 and finished 1st of June 2007, the starting data sets were sourced and, more importantly, the four NMAs defined the target data sets of the generalisation process as set of constraints.

This paper addresses some fundamental issues concerning the formalisation of cartographic constraints that were met during the preparation phase. In Section 2 we describe our proposal to organise the constraints in an extended typology, based on past research and the EuroSDR project experiences. Section 3 describes the process of defining the constraints within the EuroSDR project: how did the NMAs come to the definition of constraints; what information is used; what were the problems encountered during the formalisation of constraints. In Section 4 the results are reported of comparing the constraints of the different NMAs in order to deduce general principles for cartographic constraints formalisation of topographic data. How the constraints can be used in the evaluation of generalisation output is described in section 5. The paper ends with conclusions.

2. TYPOLOGY OF CARTOGRAPHIC CONSTRAINTS

This section describes the typology of cartographic constraints that we propose. The typology (section 2.4) is a result of past research on cartographic constraints (section 2.1, section 2.2 and section 2.3).

2.1 The usage of cartographic constraints in automated generalisation

The concept of cartographic constraints was introduced to automated generalisation by Beard (1991). The goal was to provide a framework for the flexible description of cartographic requirements in a formalised way. A first generalisation process model on the basis of constraints was developed by Ruas and Plazanet (1996). They proposed the usage of formalised constraints for conflict identification, operator selection and validation of the applied generalisation transformations, see Figure 1.

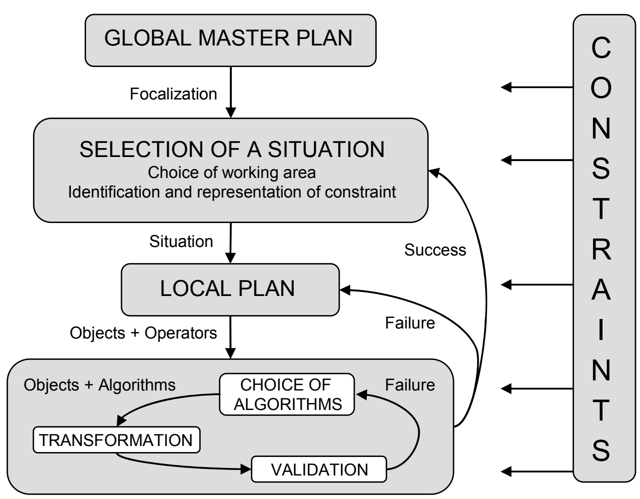


Figure 1: Generalisation process model with the utilisation of constraints (Ruas and Plazanet, 1996)

The approach from Ruas and Plazanet (1996) was inspired as they say i) by the work from Brassel and Weibel (1988), where the step of process recognition was mentioned, ii) the argumentation from Mackaness (1995) about navigation through the solution space among alternate designs and iii) the proposal from Beard (1991) with the application of constraints instead of rules.
During the AGENT[3] project (Barrault et al., 2001) the constraint-based approach was developed further, with the utilisation of micro and meso agents (Ruas, 1999) for the representation of single or group of objects. This agent types were linked with a set of constraints expressing the need for generalisation and the preservation of original characteristics. The separate formalisation of cartographic requirements with constraints offered also the possibility to evaluate generalisation *solutions*. Bard (2004) proposed a methodology to assess the quality of cartographic generalisation on the basis of three different function types. The characterisation functions capture geographic properties at the initial and the final state. The evaluation functions are used to compare both states and the aggregation functions are necessary to derive a qualitative assessment such as good, medium or bad acceptability from the quantitative values of the evaluation functions. Thus the aggregation involves the aggregation of the various assessment results to summarise data quality whereas a distinction between intra-class and inter-class aggregation is possible.

2.2 Constraint-based approaches and condition-action modelling

Another argument for using constraints in generalisation refers to the triggering and parameterisation of generalisation operators. A distinction can be made between constrained-based approaches (Ruas, 1999; Barrault et al., 2001) and condition-action modelling (Mackaness et al. 1986, Nickerson, 1988; Schylberg, 1993). In condition-action modelling also called rule-based approaches a close binding exists between case *description* on object states and the corresponding generalisation operation (*action*). Constraint-based approach gives more flexibility to the selection of generalisation operators. Thus several generalisation operations can be applied and only the best generalisation result is kept. The decision about the best suited generalisation operation is made *afterwards* based on the reduction of constraint violations.
Advantages of constraint-based approaches in contrast to condition-action modelling are the decoupling of situation description and application of generalisation operators. With that a separation of *conflict analysis* and *conflict solution* is realised. Difficulties with condition-action modelling occur when rules are defined in a contradictory way, for example i) simplify an object and ii) preserve its original shape. Constraints enable a more distinct description of mapping conflicts. If we assume for simplicity reasons only yes or no decisions of constraint violations, than with 10 constraints already 2^{10} = 1024 different situation states can be described. In case of considering 5 instead of 2 discrete constraint levels (ranging from 1-constraint is not satisfied to 5-constraint is fully satisfied), the number of cases explodes to 5^{10} = 9'765'625.
Typically in constraint-based approaches different generalisation operators or operator sequences are applied to the same situation and depending on constraint violation before and after generalisation the most promising solution are accepted. In case of iterative, step by step selection of generalisation operators, there are different strategies applicable to decide what the most promising solution is, not necessarily leading to the most optimal final result. This is due to the fact that the most promising solution may have been chosen based on high reduction of a constraint violation through one operator, which does not necessarily lead to a satisfying final solution. Possible approaches for triggering generalisation operations on the basis of constraints are simulated annealing (Ware et al., 2003), hill climbing algorithm (Regnauld, 2001) or collaborative filtering (Burghardt and Neun, 2006).

2.3 Constraint typologies

Typologies of constraints are proposed by several authors. Beard (1991) suggested a categorisation of graphic, structural, application and procedural constraints. Ruas and Plazanet (1996) refined the constraint typology on objects with legibility, shape, spatial and semantic constraints. Weibel and Dutton (1998) as well as Galanda (2003) proposed typologies of constraints that are influenced by thematic mapping. During the AGENT project a constraint typology based on five major categories were applied (AGENT, 1998) with the following explanations. *Graphic* constraints arise from feature and symbol geometry specifying basic size and proximity constraints. *Topological* constraints such as connectivity, adjacency, containment or self-intersection ensure that topology is preserved. *Structural* constraints are

[3] AGENT – Automated GEneralization – New Technology (EU research project from 1997-2000)

describing both spatial and semantic structure as well as their interdependencies. In contrast to graphic and topological constraints, structural constraints describe higher order concepts. *Gestalt* constraints relate to aesthetics and complex aspects. Finally the *procedural* constraints are defined to capture pre knowledge about the generalisation process.

2.4 Our proposal of a constraint typology
Our proposal of a constraint typology as visualised in Fig. 2 extends the former approaches.

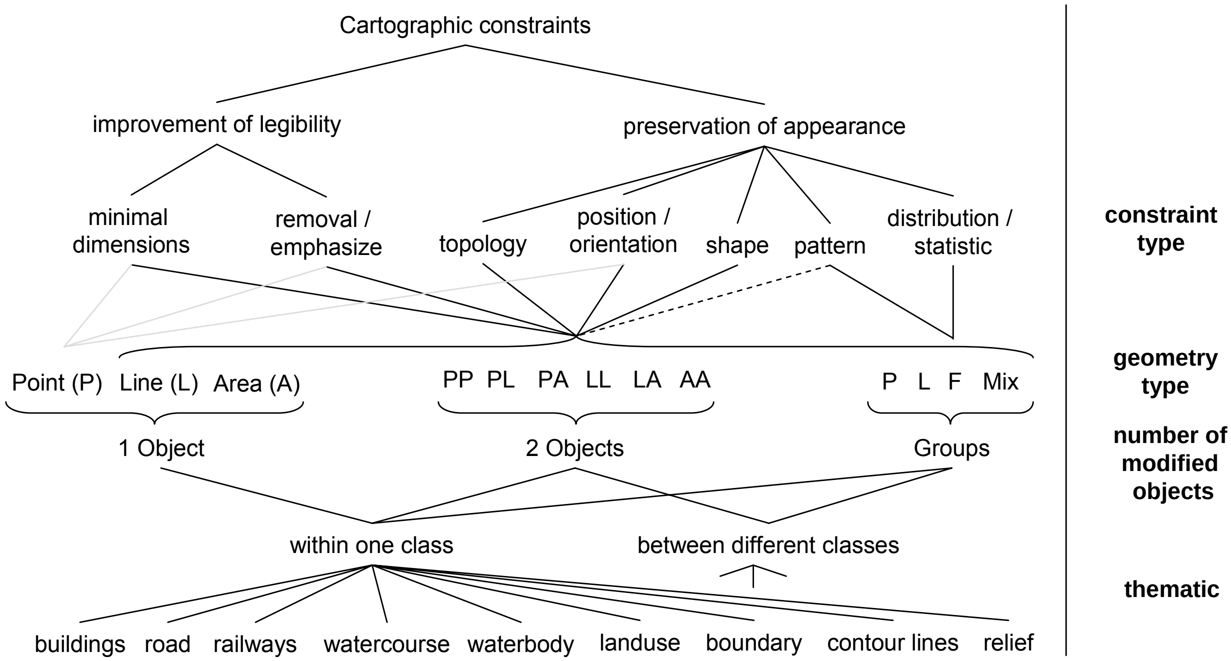


Figure 2: Typology of constraints

First of all a distinction is made between two main categories – the legibility constraints and the constraints for the preservation of appearance. The difference is that preservation constraints at the beginning of the generalisation process are completely satisfied, while legibility constraints are violated through the scale changes and the applied symbolisation. Harrie (2001) argues further legibility constraints aim at changing the data, while the preservation constraints strive to maintain them. A second major difference is that the violation of legibility constraints at the target data set can be investigated independent from the source data set. In contrast the preservation constraints have always to be calculated in correlation with the source data.
Another categorisation on the top level, which is not considered here, might be the distinction between hard and soft constraints drawing on the fact that some of the constraints have to be satisfied completely such as topological or minimal dimensions constraints, while other should be satisfied as best as possible.

On the next level a characterisation of constraint is made according to the constraint type synthesising some of the above mentioned typologies. Legibility constraints are subdivided into *minimal dimension* and *removal/emphasize* constraint type. Both contain constraints, which force the application of generalisation operations. The *minimal dimension* constraints ensure that objects or object parts are large enough to be clearly visible. The *removal/emphasize* constraints are defined for objects with less respectively more semantical importance or as consequence of minimal dimension violation or combination of both cases. The constraint type was added as consequence of the large number of mentioned cases by the constraint formalisation of NMAs (see also Section 4) despite of the fact that it represents a concrete generalisation action instead of a simple constraint.
The preservation constraints are subdivided into five further categories such as *topology, position/orientation, shape, pattern* and *distribution/statistic*. In comparison with the AGENT classification the constraint type of *topology* is identical. While cartographic requirements on translation and rotation are covered by *position/orientation* constraint type, the deformations of the object themselves are described

by *shape* constraint type. The *pattern* constraints are used to model requirements on repetitions of objects or object parts. The *distribution/statistics* constraint type finally models more global effects such as black/white ratio, trends or clusters.
The semantical aspects mentioned together with *structural* constraints in the AGENT classification can be combined with all constraint categories from our point of view, e.g. a constraint such as "the street connected with a house has to be kept" combines topology and semantic. Therefore semantical aspects are described at a thematic level relevant to all constraint categories. A major distinction can be made for objects belonging to the same thematic class in contrast to constraints relating objects of different classes.

Two further dimensions of the constraint typology are introduced with 'the number of objects to be considered' and 'the geometry type'. Only some of the constraint types have a meaning for one object of geometry type point (grey lines in Fig. 2), e.g. minimal dimensions, removal/emphasize and position constraint. Topology constraints for one object of line or area type can be found with the self-intersection constraints. Distribution/statistic constraints are only defined for groups. Pattern constraints are mainly relevant for object groups as well, but additionally patterns are imaginable on the level of objects part valid for one object for example curve pattern in mountainous area. Because the relation refers only to the object *part* this reference is shown with a dashed line.

3. THE PROCESS OF CARTOGRAPHIC CONSTRAINTS FORMALISATION

Much information on generalisation requirements needed for the formal specification of cartographic constraints is already available at NMAs either implicitly or explicitly. This information is available in data models and data structures of current datasets, in 'cookbooks' with generalisation rules to guide the cartographer, in tools in which part of the process has been automated and/or in heads of humans that interactively execute or trigger (parts of) the generalisation process. All these information sources were used to define the constraints of the target data sets within the EuroSDR project. The potentials of these sources, as well as their implications to be used in the process of cartographic constraints formalisation is discussed in this section (section 3.2 till 3.6). This discussion is based on the initial experiences in the EuroSDR project. Section 3.1 first reports on the EuroSDR project approach on constraints formalisation.

3.1 EuroSDR project approach on cartographic constraints formalisation

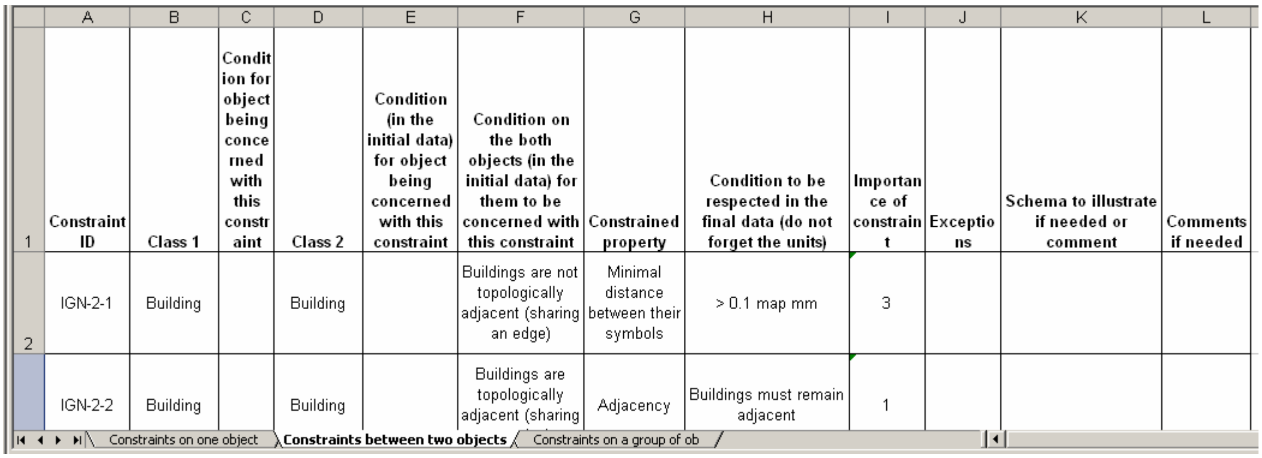
To meet the objective of the EuroSDR project to generalise complete maps, the most relevant transformations were identified required to generate smaller scale data sets from large scale data sets. These transformations were categorised by feature type. This yielded transformations classified by the following feature classes: buildings, roads, railways, hydrography network, land use (in a planar topology structure), relief, isolated lines (embankment, hedges), isolated points (poi, symbols), isolated areas and coastlines. In order to cover all these transformations four generalisation cases were selected as test cases (see Table 1).

Table 1: Overview of test cases used within the EuroSDR generalisation project

Source dataset	***Type of area***	***Target dataset (DCM)***
1:1250	Urban area	1:25k OSUK
BD Topo (~1:10k)	Mountainous area	1:50k IGN
TOP10NL	Rural area	TOP50NL
1:25k	Costal area	1:50k ICC

In order to have coherently structured sets of constraints, a template was prepared and provided by the project team (see Figure 3). The template defines constraints on one object, between two objects and on a group of objects. In a first step these templates were filled by the four NMAs for their specific cases and cartographic requirements. It should be noted that the constraints were specifically generated for the test cases of the EuroSDR project, i.e. the constraints that were generated are not meant to offer a complete description of cartographic generalisation output.

The comparison of these constraints in section 4 will show quantitative differences in constraint frequencies among the different NMAs. However there were also qualitative differences in constraint formalisation: different terms and constraints were used for similar situations. These differences resulted from the fact that the NMAs specified the constraints independently from each other based on the constraint template without further information about the degree of formalisation. Both the qualitative and the quantitative differences in constraint specification forced the project team to harmonise the constraints provided by the NMAs in order to simplify the generalisation tests and the following evaluation procedure.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Constraint ID	Class 1	Condit ion for object being conce rned with this constr aint	Class 2	Condition (in the initial data) for object being concerned with this constraint	Condition on the both objects (in the initial data) for them to be concerned with this constraint	Constrained property	Condition to be respected in the final data (do not forget the units)	Importan ce of constrain t	Exceptio ns	Schema to illustrate if needed or comment	Comments if needed
2	IGN-2-1	Building		Building		Buildings are not topologically adjacent (sharing an edge)	Minimal distance between their symbols	> 0.1 map mm	3			
	IGN-2-2	Building		Building		Buildings are topologically adjacent (sharing	Adjacency	Buildings must remain adjacent	1			

Constraints on one object / Constraints between two objects / Constraints on a group of ob

Figure 3: Screenshot of the template used for formalisation of cartographic constraints in the EuroSDR project

The process of constraint harmonisation started with extracting frequent defined constraints covering the same or similar situations. That implied a classification in two constraint categories: specific constraints (mostly applicable to only one data set) and generic constraints (applicable to all four data sets if necessary). After having extracted constraints covering the same or similar situations from the set of constraints provided by the NMAs, these constraints were standardised in a way that they are applicable to all four data sets. This resulted in a list of constraints that can be regarded as generic constraints for topographic map generalisation at mid-scale (within the project!). All four NMAs agreed with the list and they redefined their constraints as generic constraints using their own feature classes, thresholds, parameter values and also preferred actions (i.e. what operator should be applied to satisfy the constraint in case of a conflict). The preferred action was added (not available in the first template) since it was realised that this information is commonly available at NMAs and can help the generalisation process considerably (see also section 3.4). There are a few constraints that remained NMA specific constraints because they are dealing with very specific situations.
The harmonisation process, resulting in 21 generic constraints, was carried out for constraints on one object, on two objects and on group of objects. Some examples of homogenised constraints on one object are (parameter values are in map units):

- Minimal dimensions
  - *Area*: target area > x map mm^2; target area = initial area ± x %
  - *Width of any part:* target width > x map mm
  - *Area of protrusion/recess:* target area > x map mm^2
  - *Length of an edge/line:* target length > x map mm
- Shape
  - *General shape:* target shape should be similar to initial shape
  - *Squareness:* [initial value of angle = 90° (tolerance = ± x°)] target angles = 90°
  - *Elongation:* target elongation = initial elongation ± x %
- Topology
  - *Self-intersection*: [initially, no self-intersection] no self-intersection must be created
  - *Coalescence*: coalescence must be avoided
- Positon/Orientation
  - *General orientation:* target orientation = initial orientation ± x %
  - *Positional accuracy:* target absolute position = initial absolute position ± x map mm

After the process of harmonisation a higher level of formalisation of those generic constraints, including a more detailed study on possible languages, will be the next task in the EuroSDR project. This formalisation designates the transformation of a constraint into a formal description that is interpretable by computers. It is obvious that the degree of formalisation of the above-mentioned constraints varies strongly and not all constraints are interpretable by computers. For instance, most metric constraints can easily be formalised while shape constraints are insufficiently formalised. This lack of formalisation is also true for generic constraints on two objects and on group of objects.

Some examples of homogenised constraints on two objects are:

- Minimal dimensions
  - *Minimal distance:* target distance > x map mm
- Topology
  - *Connectivity:* [initially connected] target connectivity = initial connectivity
- Position
  - *Relative position:* target relative position = initial relative position

Some examples of homogenised constraints on group of objects:

- Shape
  - *Alignment: initial* alignment should be kept
- Distribution&Statistics
  - *Distribution of characteristics:* target distribution should be similar to initial distribution
  - *Density of buildings (black/white):* target density should be equal to initial density ± x %

These examples also show that further effort is required in order to be able to formalise such constraints. Partially this formalisation task has to be solved in order to be able to quantitatively evaluate constraint violations after application in different generalisation systems.

3.2 Data models

Data models define what objects and attributes should be defined in topographic data sets at various scales. By comparing data models at different scales it can be determined which objects at a larger scale should be aggregated to form objects at smaller scales. This information is important input for the constraints formalisation. Comparing data models at different scales may for example lead to the constraint (taken from the Dutch case):

type_of_land_use is in {“orchard”, “tree nursery”, “forest with deciuous trees”, “forest with evergreen trees”, “forest with deciduous and evergreen trees”, “cropland”, “fruit plantation, nursery”, “heath”, “poplars”} then instances are to be aggregated

From the experiences it can be concluded that it was not always straightforward to translate the information from the data models into constraints. It appeared for example that in some cases more information is kept at the smaller scales. Obviously this is not a problem at the time the topographic data sets are generated and produced independently. For example in TOP10NL the class ‘metro’ is identified, whereas in the TOP50NL data set (not existence in production line; only produced specifically for this test) a subdivision is made between ‘metro below the surface’ and ‘metro above the surface’. It is clear that in this case it will not be possible to derive the smaller scale information from the large scale data set. Another problem that was encountered when analysing the data models, is that data models, mostly modelled using the Unified Modelling Language, usually define very well the parent-child relationships. However contextual relationships such as “a building must always be accessible by a road” cannot directly be deduced from data models. In this specific case the information is implemented as software code to assure consistency during the data creation process. In conclusion current data models, although (some kind of) formal descriptions of data sets, need human interpretation to deduce formal cartographic constraints to be used in implementation.

3.3 Data structures

Data structures of currently available data sets dictate also requirements for generalisation that are a source for constraints formalisation. For example does the starting data set supports planar and network topology which should be remained after generalisation? Planar topology is only relevant in the TD Kadaster data set; network topology is supported in all data sets. Besides these common data structures, there are NMA specific structures dictating generalisation requirements. For example the ICC data set does not contain explicit polygons; only the boundaries of polygons are indicated and attributed. A polygon can be generated finding the closed polygons starting with the centroids of polygons that are indicated in the data set as well. Lines can only have one classification, depending on the priority of classification (e.g. city wall has higher priority than building). However lines may belong to more than one polygon object. Other specific NMA data structure is the representation of buildings: whether the geometry represents the real geometry in the field or whether the geometry represents a building symbol which size is related to the size in the field. Another example of NMA specific data structure that needs to be defined in the constraints is the multiple-geometry representations of roads (heartlines and polygons) in the Dutch TOP10NL.

3.4 Cookbooks

Many NMAs have defined generalisation rules in cookbooks to support the work of cartographers. For example the generalisation rules for the Dutch TOP50 have been laid down in a document (Topografische Dienst Kadaster, 2005). However these rules were not defined to guide fully computerised generalisation processes. Instead they are meant to be interpreted by humans. It was therefore not always easy to translate these rules into formalised constraints. Examples of such ambiguous rules are: “if the density of individual buildings is sufficiently, the buildings can be aggregated into built_up area”. When is the threshold of “sufficiently high” reached? What spatial extent is used to measure the building density? For automation, building blocks delineated by network features such as railways, roads and hydrography need to be generated. However these partitions have not always been followed by humans as can be seen in the example of Figure 4.

Another problem that was encountered when analysing the cookbooks is that some generalisation information is easier to define in a sequence of rules than in constraints. For example the actions in the rule that buildings at a distance between 10 and 7.5 meter should be *displaced* and buildings at a distance smaller than 7.5 meter should be *aggregated* (example taken from ICC) are difficult to cover by a constraint. This was solved by adding the possibility to link a preferred action when a constraint is violated. The preferred action can then be linked to a specific condition. As a consequence several actions can be linked to one constraint, dependent on the condition.

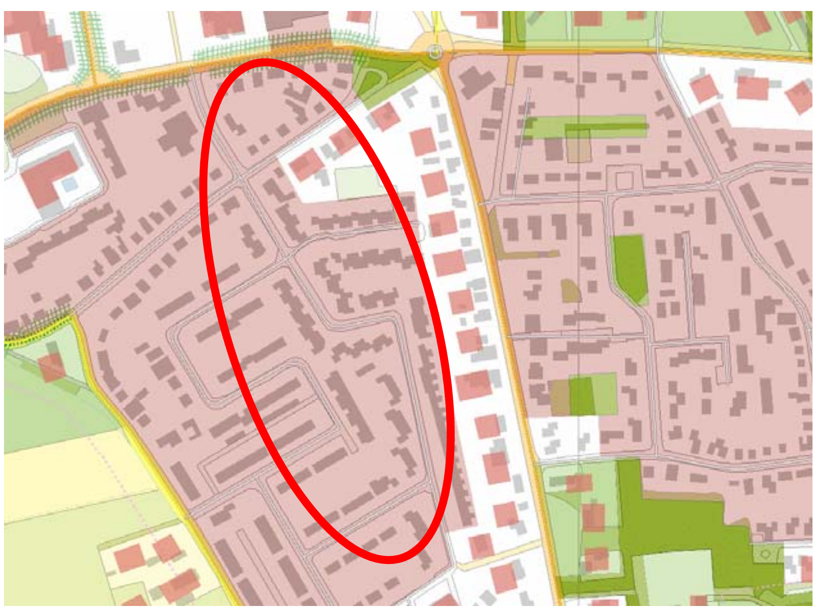


Figure 4: Top50 projected on TOP10NL, using transparency. For the aggregation of individual buildings a spatial partition was chosen based on the building pattern that requires human interpretation.

3.5 Information in program code

Constraints information might be available in programming code supporting current production lines of NMAs. For example the constraints valid for groups of objects in the IGN data set. These groups are not part of the data model, but need to be created before violation of the constraint can be detected. These are for example urban blocks (in order to keep spatial distribution pattern of buildings or to avoid a density conflict), groups of lakes (to remain the spatial distribution of lakes) and roundabouts (to remove them and to extend roads to the centroids of the roundabout). Another example of information needed for generalisation implemented in software code was already mentioned before, which are semantical relationships such as between buildings and roads.

When the software code is accessible by the generalisation, it is not a problem. However this is not always the case, e.g. when it is code programmed to build the data sets. The information captured in the code must then be exported somehow to be accessible for the generalisation process.

3.6 Information available with generalisation experts

Generalisation rules have been set up to support cartographers in the generalisation process. At the time these guidelines were defined, it was known that these guidelines were not the only information source for the generalisation process. The process is supervised by the cartographer who interprets and orchestrates the guidelines in the optimal sequence, also taking into account his/her own vision of a specific situation. For example the Dutch generalisation guidelines are accompanied by the general guideline that the cartographer should always take into account the field characteristics which (s)he can deduce from aerial photographs. In the constraint-based approach, the constraints are the only source that guides the generalisation process. Consequently all information that is currently added to the generalisation process by humans need to be formalised. In the EuroSDR project this is done by involving cartographers in the process of cartographic constraints formalisation. Since the constraints are defined before any computerised generalisation is applied, it cannot be guaranteed that all information available with the cartographer is captured in the constraints. Therefore the list with constraints originating from cartographic experts will need to be completed in an iterative process.

4. Comparison of different NMA specifications

This section analyses the first sets of cartographic constraints provided by the NMAs. In this section the constraints of IGN, ICC and TDK are analysed. The constraints of OS UK were not taken into account since these were not available at the time this analysis was carried out. The comparison presented in this section was carried out before harmonisation of the constraints took place. In fact the results of this comparison were used in the harmonisation process as described in section 3.1.

Because of the provided template, the sets of constraints do not differ according to their structure, but rather to their content. The comparison presented in this section gives a first impression of constraint specification considering similarities and differences according to several issues. The comparison of the constraints provided by the NMAs is done from a quantitative and a qualitative point of view. Following issues will be considered:

- section 4.1: general similarities and differences (type of area, degree of scale reduction, thematic classes)
- section 4.2: number of constraints and its dependency on the thematic classes, area types and spatial scope
- section 4.3: NMA constraints in relation to the proposed typology

4.1 General similarities and differences

From Table 1 it can be seen that the degree of scale reduction, as well as the starting and target scales differ from each other. The degree of scale reduction as well as the target scale affects the kind of cartographic constraints because the larger the degree of scale reduction the stronger the model and graphical simplification (AGENT Consortium, 1998). The target scales define among others the map content because of graphical and spatial limitations.

Not only the degree of scale reduction and the target scales but also the different *types of areas* lead to different constraint definitions. The four different test data sets differ from each other according to number and granularity of specified thematic classes. The ICC data set contains 46 various thematic classes whereas in the IGN and TDK data sets 5 respectively 6 thematic classes can be found. Consequently the thematic classes covered by ICC constraints are more specific compared to the thematic classes covered by IGN and TDK constraints (e.g. distinguish between more types of roads). In addition ICC provides a test data set which is a representative of a costal area meaning that certain thematic classes appear only in this data set. Furthermore the number of objects varies within the thematic classes as consequence of the chosen types of areas. While in topographic maps illustrating urban areas a large number of building features exists, only a small number of building features can be found in a topographic map illustrating a mountainous area. It should be noted that some feature classifications were adjusted (simplified) for the project by the NMAs.

4.2 Number of constraints

These differences have effect on the kind of constraints as well as on the number of constraints. The ICC set of constraints contains 151, the IGN set contains 57 and finally the TDK set contains 46 constraints. The reason that ICC has defined more constraints is related to the applied granularity of thematic classes. While IGN and TDK used simplified schema for testing purposes, the ICC worked with their normal ones. A further analysis of the number of constraints can be made according to the thematic affiliation. This kind of analysis shows the distribution of constraints to the thematic classes and enables a statement about the individual weighting and the importance of thematic classes in topographic maps.

In order to allow for a meaningful comparison, we categorise the thematic classes provided by the NMAs, especially the ones proposed by the ICC, and term them coherently. Nevertheless, not all categories have specified constraints for the test data sets of the different NMAs. Constraints within a thematic class as well as the constraints between different thematic classes are incorporated in Table 2. For instance, a constraint, which affects the building and the road class (e.g. minimum distance), is counted in both categories. Furthermore there are constraints which refer to different or any thematic classes. Thus the number of constraints in Table 2 is larger than the above mentioned numbers.

Table 2: Number of constraints according to thematic classes

	ICC	**IGN**	**TDK**	**Total**
Building	49	36	12	97
Road	35	15	21	71
Railways	Doesn't exist	Doesn't exist	4	4
Watercourse	29	6	5	40
Waterbody	26	6	2	34
Landuse/Landcover	32	6	12	50
Contour lines	16	2	Doesn't exist	18
Boundary	Doesn't exist	Doesn't exist	Doesn't exist	0
Relief	2	2	Doesn't exist	4

Considering these amounts and distribution a few observations can be made. The large amount of constraints on building themes as well as on road themes protrudes. Furthermore it is interesting that none of the NMAs compiles constraints on boundary themes. The constraints on railway theme are only dealt with by one of three NMAs because this theme is only available in one test case. It is now arguable whether the *number* and *complexity* of map objects, the extent of scientific research (e.g. generalisation of building features is frequently object of interest) or just the incomplete handling is the reason for the quantitative heterogeneity.

The distribution of constraints according to the different types of data sets (type area) is noticeable since the ratios of number of constraints within a data set differ from each other. The data set provided by the ICC illustrates a costal area and could be characterised by the immediate proximity to the sea and by the densely settlement in the coastal area. Consequently, most constraints pertain the building and hydrography themes. The data set provided by the IGN illustrates a mountainous area and could be characterised by less density of building and road themes and by the dominant role of the relief (or DTM). In this case the corresponding constraints pertain mostly the building and road themes, similar to set of constraints provided by the ICC. The TDK provides a data set which illustrates a rural area. This kind of area can be characterised by the dominant role of landuse/landcover themes. In this way, interdependency between area type and number of constraints can be seen.

A further comparison concerns the distribution of constraints according to the quantity of objects to be considered (see Table 3). The constraints template provided by the project team consists of three parts whereas the first part deals with constraints on one object (instance), the second part deals with constraints on two objects, either of the same or different classes, and the last part contains constraints on a group of objects.

Table 3: Constraints according to the number of involved objects

	ICC	IGN	TDK
Constraints on one object	98	32	23
Constraints on two objects	27	13	17
Constraints on a group of objects	30	12	2

It is obvious that most of the constraints refer to one object whereas fewer constraints are defined for two or more objects. Especially on the group level there seems to be a lack of constraints all the more because the structures (e.g. patterns, alignments) and spatial distributions of map features should be described at this level of analysis or evaluation. In the set of constraints provided by the TDK, practically every constraint on a group of objects refers to building themes. For instance, the road and watercourse themes are not included at this level of analysis although patterns do exist there (SCC, 2002; Zhang, 2004; Heinzle, 2006).

4.3 NMA constraints in relation to proposed typology

The next step of analysing the different sets of constraints provided by the NMAs shows how many constraints exist per category according to the typology proposed in Section 2.

Table 4: Number of constraints according to the constraint type

	ICC		IGN		TDK	
Minimal dimensions	79	54%	18	33%	26	43%
Removal / emphasize	26	18%	3	6%	29	48%
Topology	14	10%	9	17%	4	7%
Position / orientation	6	4%	6	11%	1	2%
Shape	6	4%	7	13%	0	0%
Pattern	10	7%	6	11%	0	0%
Distribution / statistics	5	4%	5	9%	0	0%

Table 4 shows that the number of constraints varies between constraint types as well as between specifications for the different test data sets. It is open if the different NMAs would specify the same *type of area* in a similar way or completely different. Furthermore, interpreting the (absolute and relative) frequency of constraints, it is necessary to consider the differences of the underlying data models. For instance, the ICC data set consists of a lot more feature classes than the IGN and TDK data sets. Thus, the ICC specified more constraints than the other NMAs meaning that more constraints should not stringently be equated with completeness. The following constraint example shows up the difficulty of the interpretation:

Minimal dimensions

- Vertices density: (target) vertices distance > 1 m

The ICC applied this metric constraint on 15 feature classes (e.g. River, River_connection, River_double_margin, River_centreline, River_urban_area_connection, etc.) whereas the data models provided by the IGN and TDK consist of considerably less feature classes (e.g. River, Part of water) meaning that they specified less or no constraints of this type. Nevertheless, it is important to analyse the relative frequency according to the constraint types in order to draw conclusions.

The constraint type of preserving minimal dimensions plays an important role in all specifications, showing the importance in the cartographic generalisation process. The removal constraints are applied more often by the ICC and the TDK. Especially TDK used this constraint heavily for removal of unimportant objects (which is in fact model generalisation) and less as consequence of graphical conflicts (minimal dimensions). IGN used it only once in combination with minimal dimensions (“building should not appear in DCM in case of: symbolisation ≠ ‘Town hall’ and initial area < 30 terrain meter”). ICC applied removal constraints both for model and cartographic generalisation. Worth mentioning is that NMAs did not incorporate emphasize constraints although one of the issues of cartographic generalisation is to emphasize the important and stressing the extraordinary (SCC, 2005, p. 41). Topological constraints are defined on a more general level such as preserve topological consistency and connectivity, no self intersection, keep adjacency or force flow directions to valleys.

Another interesting realisation is that the position/orientation constraints occur sparsely in the set of constraints and further, they refer only to building objects. A reason could be that buildings are isolated objects, which could be moved (manually or automatically) more easy than network or area objects with their connectivity constraints (topology). Similar to topological constraints also the shape constraints are formulated on a general level for example ‘target shape should be similar to initial one’. Noticeable there are only a few shape constraints defined by the TDK. Finally distribution and statistic constraints are defined in a similar way by all NMA in a general way e.g. “distribution of characteristics of group of buildings (shape, size, function) should be similar to initial distribution”. Therefore only a few constraints are defined for this type of constraints.

Figure 5: Illustration of frequency distribution based on the number of defined constraints

Figure 5 represents the relative importance of constraint category measured through the number of defined constraints provided by four NMAs. It involves the integration of the quantitative analysis into the proposed typology (see Figure 2: Typology of constraints) meaning that the relative frequency of constraints per analysis level is represented by the font size. The larger the font size of the category (e.g. minimal dimensions, building, etc.) the more constraints are defined for that category. Thus, the font size corresponds to the relative value while the values in the brackets represent the absolute number of constraint per category. The grey coloured lines represent the categorisation of defined constraints. For example, a “minimal dimension” constraint was defined on “one object”, namely on a “polygon” of the “building” class. Consequently, every combination is represented in this way.

The graphic representation of the analysis result gives an impression about the first steps in constraint specification within the EuroSDR project. The uneven distribution of the specified constraints according to the several categories is obvious. For example most of constraints are “minimal dimension” constraints defined for one object of line or area type. On a thematic level mainly constraints for buildings are specified in a more concrete way while hardly any constraints were specified for boundaries and railways. These strong differences in distribution can be found on every analysis level. Consequently, the question is: why these differences in distribution? We assume that there are different issues which have to be taken into account.

On the level “number of modified objects” most constraints are defined for one object whereas fewest constraints are defined for groups of objects. It is arguable that the extent of scientific research could be one of the reasons. Research in the generalisation domain has focused for a long time more on isolated objects than on interrelations of two or more objects, e.g. most generalisation operator works on single objects. There exists also a quantitative heterogeneity on the level “constraint type”. The number of constraints defined for different constraint types partially correlate with the number of modified objects. For instance, “minimal dimension” constraints can be specified for one object, two objects or group objects while pattern and distribution/statistics constraints can only be specified for group of objects. The distribution of constraints on the "thematic level" results probably from the relative importance of certain thematic classes within the four chosen topographic map examples.

At this point it must be said (again) that the set of constraints provided by the NMAs are deduced from different initial data sets. Consequently, the specifications of constraints provided by the NMAs were based on different data models and data structures which stringently lead to different solutions. For instance, NMA specific data structures dictate the generalisation requirements (e.g. the ICC data set does not contain explicit polygons; only the boundaries of polygons are indicated and attributed), which has an affect on the number of constraints. Summarising, the comparison of the constraints defined by the different NMAs shows their priorities of constraint specification: what do they consider as the most important generalisation requirements; what do they consider as most complicated generalisation transformation etc.

5. Evaluation of Generalisation solutions

In order to yield satisfying generalisation solutions the constraint-based approach allows for two applicabilities. On the one hand the usage of constraints during the process enables the optimal orchestration of generalisation operations, on the other hand the target data sets defined as a set of constraints offer objective and quantitative evaluation possibilities of the output. Thus, the evaluation of constraints shows whether the map is cartographically satisfactory or at least if the cartographic presentation improved during the generalisation process. The assessment of quality of generalisation solutions is therefore based on the evaluation of these two approaches (evaluation of the process and evaluation of the output).

The corresponding quality assessment of generalisation outputs can be done with following evaluation procedures:

- *Evaluation I* of generalised data sets which is based on the target data set defined as a set of constraints, the results and partly on the source data set. It compares the individual outcomes to specifications of the target data sets in order to answer: How many and which constraints are solved or violated? How satisfying was the process?

- *Evaluation II* compares the results of different systems (i.e. solutions) in order to get insight into the generalisation process: Can generalities be discovered? How different are the processes/systems? Are the same generalisation strategies (order of operations, used algorithms, values of parameters) followed by different systems? The evaluation of the process expects the usage of log files produced by the different systems in combination with templates to interactively record the actions performed by the user.

A further evaluation procedure, which can be seen as supplement to Evaluation I & II, incorporates the cartographers. It mainly involves a qualitative evaluation of the generalised data.

- *Evaluation III* asks cartographers to assess the final results.

In Figure 6 the three evaluation procedures and what aspects they compare are schematised. Evaluation I relates to the generalisation results, to the specifications of the target data and partly to the source data. It is the examination whether the cartographic constraints are solved or violated. Thereby we distinguish between two different approaches according to the proposed typology. Evaluation of legibility constraints is possible independently from the source data set. In contrast, preservation constraints can only be evaluated through the consideration of source and target data sets. Therefore linkages between source and target objects are necessary, which could be derived from IDs. But in case of aggregation, collapse and typification, the identification on the basis of IDs is impossible. For geometry type change and removal of objects it might be difficult. A simple approach is the manual assignment of corresponding objects in source and target data set. A more advanced solution could be based on automated matching operations. Ideally the vertical relations would be constructed during the generalisation process and could be exported for further analysis (Bobzien et al., 2007).

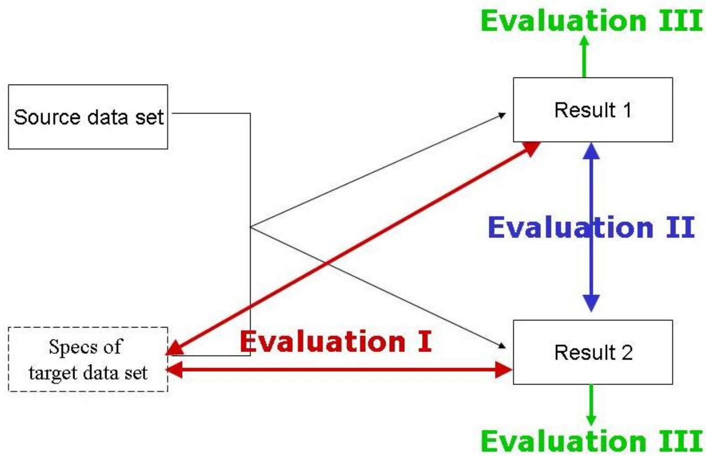


Figure 6: Evaluation procedures of generalisation solutions

The development of an automated evaluation is one of the project objectives. The implementation of the evaluation of the several types of constraints as mentioned in Section 2 will be differently difficult since the degree of complexity varies depending on the constraints. Together with the minimal dimension constraints position and orientation constraints seem to be the simplest to implement. Certainly, there are exceptions which are not equally simple to realise such as the minimal distance between two lines. The evaluation of topological constraints has the advantage that there are clear definitions exist on topological relations (Egenhofer and Franzosa, 1991; Egenhofer and Herring, 1991), but implementations which identify differences of topological relations in different data sets will take some effort. At this time it is difficult to estimate the effort necessary for the evaluation of distribution and statistical constraints. The reason is that on one side there exist a complete theory on statistics, which offers a lot of methods for calculation, on the other side it is not obvious how much distribution and statistical values can vary until cartographic requirements are violated. Finally, the implementation of pattern constraints seems to be the most difficult, therefore this part of the evaluation procedure might be solved manually.
Evaluation II involves a comparison between different results relating to source and target data set. This assessment procedure enables the comparison of different results and thus, it is also a comparison of different systems. Finally, the Evaluation III is a global and free analysis by cartographers enlightening the main quality of the proposed result, legibility, proximity to reference data set, geographical meaning preservation and detected errors. This evaluation will be based on the qualitative evaluation procedure of the AGENT prototype by cartographers and allows for a concluding evaluation of the generalisation solutions. This evaluation is based on a structured interview.

6. Conclusion

In past generalisation research it is stated that the formalisation of cartographic constraints helps to design the best generalisation scenario and that the formalisation of cartographic output provides an objective manner to evaluate the generalisation output. In Section 2 we proposed a typology of constraints that helps to structure the generalisation process in such a way that the best generalisation scenario can be designed. Within the EuroSDR project cartographic constraints have been recognised as main input for automatic generalisation and therefore four NMAs have formally specified their cartographic requirements of generalisation output for the project. The process of translating generalisation information available at NMAs into cartographic constraints as described in Section 3, teaches us how to translate all implicit and explicit generalisation information available at NMAs into formal cartographic constraints. From this section we can conclude that many information sources are available to deduce the formalised cartographic constraints. However most of this information needs to be interpreted first by humans before it can be translated into constraints. This is not a straightforward process. In addition it was concluded that not all information can be easily formalised in constraints. Section 3 also reports on the effort executed by the project team to harmonise the constraints of the four NMAs. This resulted in a list of constraints that can be regarded as generic constraints for topographic map generalisation at mid-scale (within the project!).

Section 4 compared the constraint specification from different NMAs in order to see commonness and differences. A quantitative analysis was presented embedded in the proposed constraint typology. The heterogeneous distribution of constraint specification is noticeable with priorities on minimal dimensions for isolated objects, mainly on building, roads and water themes. Finally section 5 elaborated on how the formalised constraints can be used in the evaluation of both generalisation processes and generalisation output.
Main question in the research reported in this paper remains whether and how it is possible to make all information needed for the generalisation process explicit in formal constraints. This question will be answered by experiences of the testing (which has started in June 2007) as well as the evaluation of the cartographic test outputs. As was mentioned in this paper more effort is needed to better (i.e. more) formalise some constraints in order to be interpretable by computer evaluation and to provide objective measures for quality assessment. This will be the next step in the research on cartographic constraints formalisation.

ACKNOWLEDGEMENT

We would like to thank all other members of the project team of the EuroSDR project: Cécile Duchêne and Guillaume Touya (IGN, France); Blanca Baella and Maria Pla (ICC, Catalonia); Peter Rosenstand (KMS, Denmark); Nicolas Regnauld (Ordnance Survey, United Kingdom); Maarten Storm, Annemarie Dortland and Harry Uitermark (TD Kadaster, the Netherlands), Javier Gonzalez, Magali Valdeperez and Francisco Dávila (IGN, Spain) and Karl-Heinrich Anders (University of Hannover).
They all contributed to the discussions that led to the template for the formalisation of cartographic constraints. In addition the formalised and harmonised constraints as presented in this paper are the result of activities executed by the project team.

REFERENCES

AGENT, 1998, Constraint Analysis. Deliverable A2, http://agent.ign.fr/deliverable/DA2.html

Bard, S., 2004. Quality Assessment of Cartographic Generalisation. Transaction in GIS, Vol. 8, No. 1, pp. 63-81.

Barrault, M., Regnauld, N., Duchene, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A. and R. Weibel, 2001. Integrating multi-agent, object-oriented, and algorithmic techniques for improved automated map generalization. Proceedings 20th International Cartographic Conference, Beijing, China, 6-10 August, pp. 2210-2216.

Beard, M. K., 1991. Constraints on Rule Formation. In Buttenfield, B. P., and R. B. McMaster (eds), Map Generalization: Making Rules for Knowledge Representation, Longman Group, pp. 121-135.

Bobzien, M., D. Burghardt, I. Petzold, M. Neun and R. Weibel, 2007. Multi-Representation Databases with Explicitly Modeled Horizontal, Vertical and Update Relations. Cartography and Geographic Information Science (in press).

Burghardt, D. and M. Neun, 2006. Automated sequencing of generalisation services based on collaborative filtering. In: M. Raubal, H. J. Miller, A. U. Frank and M. Goodchild (eds): Geographic Information Science. 4th Int. Conf., GIScience 2006, IfGIprints 28, pp. 41-46.

Brassel, K. E., and R. Weibel, 1988. A Review and Conceptual Framework of Automated Map Generalization. International Journal of Geographical Information Systems, Vol. 2, No. 3, pp. 229-244.

Egenhofer, M. and R. Franzosa, 1991. Point-Set Spatial Topological Relations. International Journal of Geographical Information Systems Vol. 5, No. 2, pp. 161-174.

Egenhofer, M. and J. Herring, 1991. Categorizing binary topological relationships between regions, lines, and points in geographic databases. Technical Report, Department of Surveying Engineering, University of Maine, Orono.

EuroSDR, 2007, see http://plone.itc.nl/eurosdrgen

Galanda M., 2003. Modelling constraints for polygon generalization. Proceedings 5th ICA Workshop on Progress in Automated Map Generalization, Paris, France, CD-ROM.

Harrie, L., 2001. An Optimisation Approach to Cartographic Generalisation. Ph.D. thesis, Department of Technology and Society, Lund University.

Heinzle, F., Anders, K.-H. and M. Sester, 2006. Pattern Recognition in Road Networks on the Example of Circular Road Detection, in: Raubal, M., H. J. Miller, A. U. Frank, M. F. Goodchild (Eds.): Geographic Information Science, GIScience 2006, Münster, Germany, LNCS 4197, pp. 253-267.

Mackaness, W. A., Fisher, P. F. and G. G. Wilkinson, 1986. Towards a cartographic expert system. Proceedings Auto-Carto London, Vol. 1, 578-587.

Mackaness, W. A., 1995. A Constraint Based Approach to Human Computer Interaction in Automated Cartography. Proceedings of the 17th International Cartographic Conference, Barcelona, pp. 1423-1432.

Nickerson, B. G., 1988. Automated Cartographic Generalization for Linear Features. Cartographica, Vol. 25, No. 3, pp.15-66.

Regnauld N., 2001. Constraint based mechanism to achieve automatic generalization using agent model, Proceedings GIS Research UK GISRUK, University of Glamorgan, pp. 329-332.

Ruas, A. and C. Plazanet, 1996. Strategies for Automated Generalization. Proceedings of the 7th Spatial Data Handling Symposium, Delft, the Netherlands, pp. 319-336.

Ruas, A., 1999. Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Doctoral Thesis, Université de Marne-la-Vallée.

Schylberg, L., 1993. Computational Methods for Generalization of Cartographic Data in a Raster Environment, Doctoral Thesis, Department of Geodesy and Photogrammetry, Royal Institute of Technology, Stockholm, TRITA-FMI Report 1993:7.

SSC - Swiss Society of Cartography (ed.). 2005. Topographic Maps - Map Graphic and Generalization. Swiss Society of Cartography, Publikations Series, Nr. 17. Wabern.

Topografische Dienst Kadaster, 2005, Generalisatievoorschriften TOP50vector, Handleiding

Weibel, R., and G. Dutton, 1998. Constraint-Based Automated Map Generalization. Proceedings of the 8th Spatial Data Handling Symposium, Vancouver, pp. 214-224.

Ware, J. M., C. B. Jones and N. Thomas, 2003. Automated Map Generalization with Multiple Operators: A Simulated Annealing Approach. International Journal of Geographical Information Science, Vol. 17, No. 8, pp. 743-769.

ZHANG, Q., 2004, Modeling Structure and Patterns in Road Network Generalization. ICA Workshop on Generalisation and Multiple Representation, Leicester, UK.